A Normative Model of Attention: Modulation of Neural Response

SANTIAGO JARAMILLO & BARAK A. PEARLMUTTER

Hamilton Institute, NUI Maynooth, Ireland

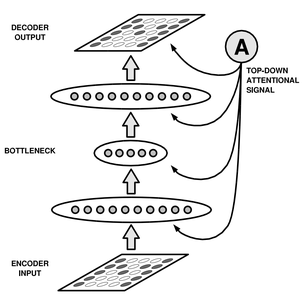
Abstract

When a sensory stimulus is encoded in a lossy fashion for efficient transmission, there are necessarily tradeoffs between the represented fidelity of various aspects of the input pattern. In the model of attention presented here, a top-down signal informs the encoder of these tradeoffs. Given an ensemble of input patterns and tradeoff requirements, our system can learn to encode its inputs optimally. This general model is instantiated in a simple network: an autoencoder with a bottleneck, innervated by a top-down attentional signal, trained using backpropagation. The only information the encoder receives concerning the semantics of the top-down attentional signal is from the optimization criterion, which penalizes the system more heavily for errors made near a simple attentional spotlight. The modulation of neural activity learned by this model qualitatively matches that measured in animals during covert visual attention tasks.

Integrating Attention & Coding

This theory of top-down attentional modulation builds upon optimal sensory encoding [1] by adding a top-down signal correlated with changing tradeoffs between coding fidelity of various features. Optimal codes can often be found analytically, but learning algorithms provide a more general approach which can accommodate the top-down attentional signal by modulating the objective function. Here we built a very simple model in order to account for a particular well-controlled experimental phenomenon involving spatial attention in the visual modality [2]. However, by avoiding analytical methods for finding the optimal code, the model can be extended to other sorts of tradeoffs in coding fidelity, such as pop-out, non-spatial attention, and non-visual modalities.

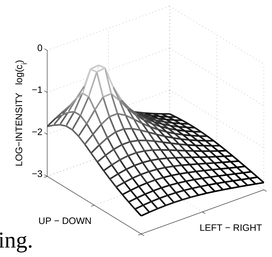
Architecture



- Auto-associative with bottleneck.
- 256:20:10:20:256
- **INPUT** 16×16 pixels and attentional signal.
- **OUTPUT** 16×16 pixels.
- Layers fully connected in a feedforward fashion.
- Additional attentional input to each layer.
- Hyperbolic tangent activation functions.
- Decoder used for optimization—not part of the brain!

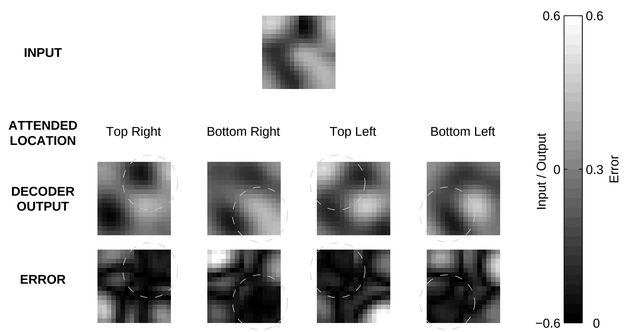
Training

- Trained as a single system (encoder/decoder).
- Backpropagation with error measure $E = \sum_i c_i (y_i - d_i)^2$
  c_i : intensity of the attentional "spotlight" at location i.
  y_i : output at location i.
  d_i : pixel i target output; $d_i = \text{input}_i$.
- Error weighting
  $c_i = 1/(1 + k^2||i - a||^2)$
  i : position on the plane.
  a : attentional input (center of the "spotlight").
  k : width parameter.



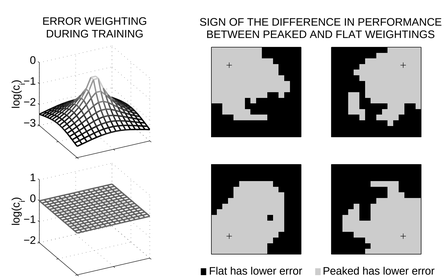
- All connections plastic during learning.
- 2,000 image (low freq. colored noise) training set.
- Center of attention uniformly distributed.
- Gaussian noise added to bottleneck units during training.

Results: Reconstruction



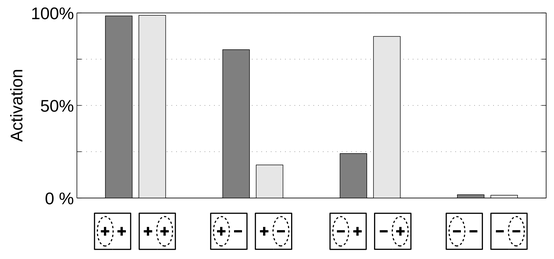
Reconstruction. One input pattern, four different attentional states as indicated by the dashed circles. The error is the absolute intensity difference between input and output.

The error inside the dashed circles is smaller than outside, *i.e.* the reconstruction is better in attended locations.



Comparison with flat error weighting. The performance difference shows that resources are redistributed.

Results: Modulation of Activity

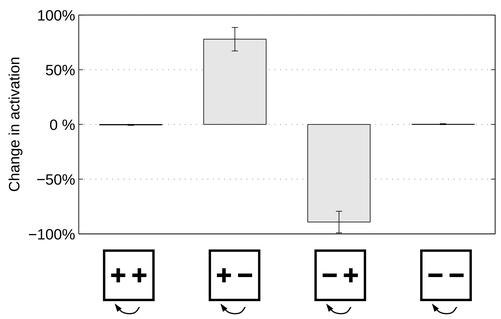


Activation of one bottleneck unit when attention is directed to left or right. Stimuli were created by combining halves of excitatory/inhibitory stimuli, as indicated by +/–. The dashed ellipse indicates the attended location.

Excitatory and inhibitory stimuli (with respect to the activation of each unit in the bottleneck) were found using the reverse correlation technique.

Clear modulation of the activation is observed depending on the top-down attentional signal.

Stimulus changes in unattended locations produce smaller modulation in activation than changes in attended locations.



Attentional modulation of unit activation. Average (over bottleneck units) change in activation when attention is shifted from right to left. Standard error also shown.

The stimuli consist of combinations of excitatory and inhibitory halves as indicated by the + and – signs. Note that the stimuli are different depending on the unit being evaluated.

All units show consistent modulation of their activity depending on the attentional state.

Discussion

The now-dominant account of low-level sensory processing posits that the nervous system encodes the sensory stimuli so as to produce internal representations which are optimal according to an appropriate information-theoretic measure. Detailed theories in this class have accounted for various previously mysterious properties of receptive fields [3]. One property of these models is that information theory takes no account of semantics, and consequently there are often many optimal codes, differing in which features of the sensory data are represented and which are discarded. This symmetry is broken in an ad-hoc fashion, by using a lower bound on the actual efficiency of the code. For instance, in a visual modality this lower bound might consist of reconstruction fidelity, with errors in each pixel weighted identically.

We instead break this symmetry by providing a top-down "semantics" signal which allows the encoder to change its representation so as to choose a code which is optimal not only in information-theoretic terms, but also in allocation of representational resources to features of current importance. This top-down signal changes with time, and the representation of the same input will in general change with it, giving rise to top-down attentional modulation of internal features.

Conclusions

- Attentional modulation integrated into information-theoretic account of receptive fields.
- Special architectural treatment of top-down modulatory signals is unnecessary.
- Effect of top-down attentional modulation can be learned.
- Modulation matches that observed in animal experiments.
- Applicable to pop-up, non-spatial attention, and across modalities.

Acknowledgements

Supported by US NSF CAREER 97-02-311, an equipment grant from Intel corporation, a gift from the NEC Research Institute, and Science Foundation Ireland grant 00/PI.1/C067.

References

[1] H. B. Barlow. Possible principles underlying the transformation of sensory messages. In W. A. Rosenblith, editor, *Sensory Communication*, pages 217–234. M.I.T. Press, 1961.

[2] S. Treue and J. H. Maunsell. Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. *J Neuroscience*, 19(17):7591–7602, 1999.

[3] J. J. Atick and A. N. Redlich. Towards a theory of early visual processing. *Neural Computation*, 2(3):308–320, 1990.

[4] S. Jaramillo and B. A. Pearlmutter. A normative model of attention: Receptive fi eld modulation. *Neurocomputing*, In press, 2003.